

AUTOMATED SPEECH RECOGNITION

BACKGROUND

[0001] Some computer systems may be adapted to detect and recognize spoken words. Typically, an input device, such as a microphone or a telephone, receives the spoken words and converts the words into an analog or digital computer readable representation. An automated speech recognition (ASR) engine may utilize the representation to detect and recognize the words.

[0002] In many situations, the ASR engine may be licensed to an organization from an external developer of the engine. The license may specify the maximum number of simultaneous connections allowed to be established with the ASR engine. Unfortunately, the number of connections needed may exceed the number of connections allowed by the license. In addition, modifying the license to increase the number of allowable connections may result in a fee imposed by the developer.

BRIEF SUMMARY

[0003] In accordance with at least some embodiments, a system comprises a first speech recognition engine, a second speech recognition engine, and evaluation logic coupled to the first and second speech recognition engines. The evaluation logic evaluates the first and second speech recognition engines based on evaluation voice signals from a user and, based on the evaluation, selects one of said speech recognition engines to process additional speech signals from the user.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] For a detailed description of exemplary embodiments of the invention, reference will now be made to the accompanying drawings in which:

[0005] Figure 1 shows a system constructed in accordance with embodiments of the invention and including a speech recognition module;

[0006] Figure 2 shows a block diagram of the speech recognition module of Figure 1; and

[0007] Figure 3 illustrates a flow chart of an exemplary connection procedure in accordance with embodiments of the invention.

NOTATION AND NOMENCLATURE

[0008] Certain terms are used throughout the following description and claims to refer to particular system components. As one skilled in the art will appreciate, various companies may refer to a component by different names. This document does not intend to distinguish between components that differ in name but not function. In the following discussion and in the claims, the terms “including” and “comprising” are used in an open-ended fashion, and thus should be interpreted to mean “including, but not limited to... .” Also, the term “couple” or “couples” is intended to mean either an indirect or direct electrical connection. Thus, if a first device couples to a second device, that connection may be through a direct electrical connection, or through an indirect electrical connection via other devices and connections.

DETAILED DESCRIPTION

[0009] The following discussion is directed to various embodiments of the invention. Although one or more of these embodiments may be preferred, the embodiments disclosed should not be interpreted, or otherwise used, as limiting the scope of the disclosure, including the claims. In addition, one skilled in the art will understand that the following description has broad application, and the discussion of any embodiment is meant only to be exemplary of that embodiment, and not intended to intimate that the scope of the disclosure, including the claims, is limited to that embodiment.

[0010] Figure 1 shows an automated speech recognition (ASR) system 100 configured in accordance with embodiments of the invention. As shown, system 100 comprises a computer system 102, a network 104, and one or more audio devices 106. The computer system 102 comprises a central processing unit (CPU) 108, a memory 110, and an input / output (I/O) interface 112. The

memory 110 may comprise any type of volatile or non-volatile memory, such as, by way of example only, random access memory (RAM), read-only memory (ROM), or a hard drive. Stored within the memory 110 are one or more speech recognition (SR) modules 114.

[0011] The network 104 couples together the audio device 106 and the computer system 102 and facilitates the exchange of data between the audio device 106 and the computer system 102. The audio device 106 may comprise a telephone, and the network 104 may comprise the infrastructure of telephone lines and signal switches that route telephone calls. In some embodiments of the invention, the network 104 may be an internet protocol (IP) network, such as the Internet, and the audio device 106 may comprise a voice-over-IP (VoIP) transmitter and receiver.

[0012] The I/O interface 112 couples together the network 104 and the computer system 102 and facilitates the exchange of data between the network 104 and the computer system 102. The I/O interface 112 comprises hardware that is capable of establishing a connection with the network 104, such as modems and network adapters. "Utterances" from a user 116 of the audio device 106 may be converted into an analog or digital representation by the audio device 106 and routed through the network 104 to the I/O interface 112. As used herein, an utterance is a vocalization that represents a certain meaning to the system 100. Utterances may be a single word, a few words, a sentence, or even multiple sentences. Once received by the I/O interface 112, the representation may be stored in the memory 110 and processed by the SR module 114 and the CPU 108.

[0013] Figure 2 shows an exemplary implementation of the SR module 114 in greater detail. As shown, the SR module 114 comprises an interactive voice response (IVR) platform 202, a dialog manager 204, an ASR switch 206, a port monitor 208, an evaluator 210, a primary ASR engine 212, and one or more secondary ASR engines 214. One or more interfaces 216, 218, and 220 may facilitate the transfer of data and control signals between components of the SR module 114 via a standard protocol, such as Media Resource Control Protocol (MRCP). The SR module 114 may be implemented via software that is executed

by the CPU 108 (Figure 1) or via a combination of software and hardware. Although the SR module 114 is shown as residing in the single computer system 102 (Figure 1), the SR module 114 may be distributed to a plurality of distinct computer systems that are coupled together via the network 104 or another connection means.

[0014] The IVR platform 202 may comprise a plurality of speech recognition applications that facilitate messaging, portals, and other enhanced voice-enabled interactive services. Typically, the IVR platform 202 is capable of handling a plurality of simultaneous user sessions. Each user session represents an established connection between the IVR platform 202 and the user 116 of the system 100.

[0015] To enable ASR functionality, the IVR platform 202 may establish connections with the primary and secondary ASR engines 212 and 214 through the dialog manager 204. The interface 216 negotiates the desired connections with the ASR switch 206. The ASR switch 206 may establish and release connections to the primary ASR engine via the interface 218 and establish and release connections to the secondary ASR engine 214 via the interface 220.

[0016] The primary and secondary ASR engines 212 and 214 may comprise logic that performs ASR functions, such as signal processing and matching. The logic embodied in the ASR engines 212 and 214 may be the same or different from each other. If ASR logic is different in the engines 212 and 214, the resulting relative accuracy or performance of the engines may differ. The primary and secondary ASR engines 212 and 214 may be representative of a commercial grade ASR engine and an in-house or open source ASR engine, respectively.

[0017] The primary ASR engine 212 is used pursuant to an associated license that specifies the number of simultaneous connections that may be established between the IVR platform 202 and the primary ASR engine 212. The license may carry an associated fee that increases with the larger numbers of licensed connections. For example, a twenty-connection license may cost twice the amount of a ten-connection license. The secondary ASR engine 214 may not have an associated license and thus may establish any number of connections

with the IVR platform 202. The secondary ASR engine 214 may be exemplary of an open source ASR engine.

[0018] The embodiments of the invention effectively reduce the number of connections established to the primary ASR engine 212 by utilizing the secondary ASR engine 214 whenever a predetermined evaluation condition is met. Since the secondary ASR engine 214 may not have an associated licensing fee, the overall costs associated with ASR functionality in the system 100 may be reduced.

[0019] Figure 3 shows a flow chart of an exemplary ASR connection procedure in accordance with embodiments of the invention should be reviewed with Figure 2. The dialog manager 204 may initiate the procedure when the user 116 attempts to utilize the ASR system 100 (block 302). In block 304 connections may be established between the IVR platform 202 and both the primary and secondary ASR engines 212 and 214 by the ASR switch 206. Both ASR engines 212 and 214 are invoked (block 306), and an evaluation set of utterances from the user 116 may be evaluated (block 308) by the evaluator 210. The evaluation set of utterances may comprise the first n (e.g., 5) words spoken by the user 116. Based on the evaluation (described below), the primary ASR engine 212 or the secondary ASR engine 214 is selected to process the user's future utterances within the same session. If the primary ASR engine 212 is selected, the connection to the secondary ASR engine 214 is released (block 310). After the user's session completes, the primary ASR engine 212 may be released (block 312). If the secondary ASR engine is selected during the evaluation, the primary ASR engine 212 is released (block 314), and the secondary ASR engine 214 may continue to process the user's utterances. The connection to the secondary ASR engine 214 may be released after the user's session completes (block 316). If neither the primary ASR engine 212 nor the secondary ASR engine 214 pass the evaluation criteria, the ASR switch 206 may be configured to optionally fallback to an alternative communications mechanism, such as Dual Tone Multi-Frequency (DTMF) (block 318). The alternative communications mechanism utilizes a non-ASR input mechanism, such as the touch tone frequencies associated the button the user has pressed. Thus, before validation both the primary and secondary

ASR engines 212 and 214 handle the user's session. After validation the user's session is solely handled by the first ASR engine 212, the second ASR engine 214, or optionally by the fallback mechanism 318.

[0020] Referring again to Figure 2, the evaluator 210 may use evaluation criteria to determine whether the primary ASR engine 212, the second ASR engine 214, or optionally the fallback mechanism will handle the user's session after evaluation. The evaluation criteria may be verification-based, response time-based, confidence-based, continuation-based, or a combination thereof. In addition, the number of utterances n used for the evaluation may be decided by a static analysis of the dialog structure associated with the IVR platform 202, a dynamic assessment based on preceding utterances, or a combination thereof.

[0021] Verification-based evaluation criteria compare the output of the primary and secondary ASR engines 212 and 214. If the secondary engine 214 produces output identical to the primary ASR engine 212, the secondary ASR engine 214 may be used, thereby allowing other connections to use the licensed ports of the primary ASR engine 212.

[0022] Response time-based evaluation criteria determine (e.g., measure), a parameter such as the response time of the primary and secondary ASR engines 212 and 214. If, compared to the primary ASR engine 212, the secondary ASR engine 214 has an identical or shorter response time, the secondary ASR engine 214 may be used after validation.

[0023] Confidence-based evaluation criteria use a confidence score generated by the primary and secondary ASR engines 212 and 214 during the evaluation. A threshold may be set that determines when the evaluator 210 should select the secondary ASR engine 214 over the primary ASR engine 212. For example, the threshold may represent a fraction of the confidence score obtained from the primary ASR engine 212. If the confidence score of the secondary ASR engine 214 is equal to or higher than the threshold level, the secondary ASR engine 214 may be utilized.

[0024] Continuation-based evaluation criteria determine whether a user has successfully navigated through an ASR menu. For example, if the user is able to reach a menu beyond the first level of a menu system with both ASR engines 212

and 214, the secondary engine 214 may be selected and utilized for the user's future utterances. Successful navigation to a secondary level of the menu system may provide a relative indicator that the secondary ASR engine 214 is detecting and recognizing the user's voice commands.

[0025] The ASR switch 206 may use the results of the evaluation, as well as the optional port monitor 208, to determine which connections may be maintained and which connections may be released. In some embodiments, the port monitor 208 may be included and used to monitor currently used ports of the primary ASR engine 212. The port monitor 208, optionally in conjunction with the evaluator 210, determines whether the primary ASR engine 212 should be used without further consideration or whether the exemplary procedure of Figure 3 should be used to handle a user's session. For example, if the number of available ports exceeds a defined threshold, the primary ASR engine 212 may be used. If the number of available ports falls below the threshold, the procedure of Figure 3 may be used. The port monitor 208 may provide the number of currently active ports to the evaluator 210 for the evaluator 210 to determine whether the primary engine is to be used or whether the procedure of Figure 3 is to be used. Alternatively, the port monitor 208 may set a flag, send a message or assert a signal to the evaluator 210 to indicate whether the primary ASR engine 212 is to be used or whether the procedure of Figure 3 is to be used.

[0026] The above discussion is meant to be illustrative of the principles and various embodiments of the present invention. Numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.